Symmetric Pattern Based Word Embeddings for Improved Word Similarity Prediction

Roy Schwartz[+], Roi Reichart [*] and Ari Rappoport[+]

[+]The Hebrew University, [*]Technion IIT

CoNLL 2015

Apples and Oranges

fruits

sweet

juicy

round

Apples and Oranges

fruits

sweet

juicy

round

Apples $\boxed{\text{and}}$ Oranges

fruits

sweet

juicy

round

X $\boxed{\text{and}}$ Y

fruits

sweet

juicy

round

X and Y

Symmetric Patterns

Overview

- **The problem**
  - *Word embeddings do not capture pure word **similarity***
- **The Solution**
  - ***symmetric patterns**-based word embeddings*
  - *First embeddings to support for **antonyms** (e.g., good/bad) w/o using a dictionary*
- **Results**
  - ***5.5%** improvement over six state-of-the-art models*
  - ***10%** improvement with a **joint** model*
  - ***20%** improvement on **verbs***

Word Similarity

- Whether two words are **semantically** similar
  - cats are similar to dogs

Word Similarity

- Whether two words are **semantically** similar
  - cats are similar to dogs
- Definition is not entirely clear
  - Synonyms (i.e., share the same meaning)
  - Co-hyponyms (i.e., belong to the same category)

Word Similarity

- Whether two words are **semantically** similar
  - cats are similar to dogs
- Definition is not entirely clear
  - Synonyms (i.e., share the same meaning)
  - Co-hyponyms (i.e., belong to the same category)
- Human judgment evaluation

Vector Space Models

DS Hypothesis (Harris, 1954)

... tokens to date, **friend** lists and recent ...

... by my dear **friend** and companion, Fritz von ...

... even have a **friend** who never fails ...

... by my worthy **friend** Doctor Haygarth of ...

... and as a **friend** pointed out to ...

... partner, in-laws, relatives or **friends** speak a different ...

... petition to a **friend** Go to the ...

... otherwise, to a **friend** or family member ...

...images from my **friend** Rory though - ...

... great, and a **friend** as well as a colleague, who, ...

...

Examples taken from the ukwac corpus (Baroni et al., 2009)

Vector Space Models

DS Hypothesis (Harris, 1954)

... tokens to date, **friend** lists and recent ...

... by my dear **friend** and companion, Fritz von ...

... even have a **friend** who never fails ...

... by my worthy **friend** Doctor Haygarth of ...

... and as a **friend** pointed out to ...

... partner, in-laws, relatives or **friends** speak a different ...

... petition to a **friend** Go to the ...

... otherwise, to a **friend** or family member ...

.. images from my **friend** Rory though - ...

... great, and a **friend** as well as a colleague, who, ...

...

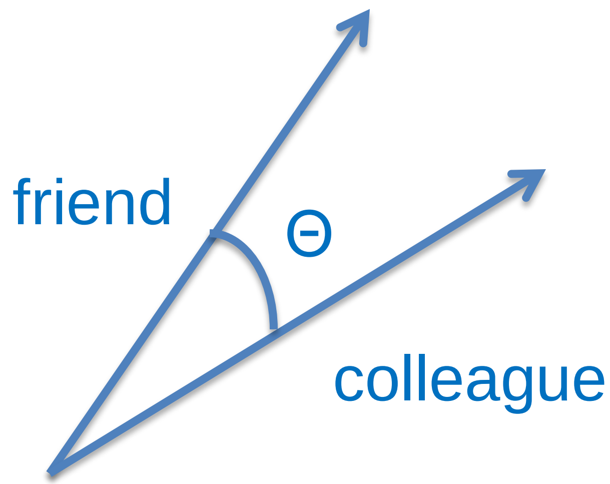
Examples taken from the ukwac corpus (Baroni et al., 2009)

Vector Space Models

$$\begin{pmatrix} 0 \\ 0.5 \\ 0.76 \\ -0.12 \\ 0.76 \\ 0 \\ 0 \\ -0.51 \\ . \\ . \\ . \end{pmatrix}$$

Vector Space Models

$$\begin{pmatrix} 0 \\ 0.5 \\ 0.76 \\ -0.12 \\ 0.76 \\ 0 \\ 0 \\ -0.51 \\ . \\ . \\ . \end{pmatrix}$$



Similarity or Relatedness?

Hill et al., 2014

cup coffee

Similarity or Relatedness?

Hill et al., 2014

sugar

milk

cup

coffee

tea

hot_water

cookie

Similarity or **Dis**similarity?

tall **short**

Similarity or Dissimilarity?

cm

feet

tall

short

John

is

length

*Current Vector Space Models do not Capture (**pure**) Word **Similarity***

Symmetric Patterns Contexts

Davidov and Rappoport, 2006

from **X** to **Y**

X and **Y**

X or **Y**

neither **X** nor **Y**

X as well as **Y**

Symmetric Patterns Contexts

Davidov and Rappoport, 2006

bright and **shiny**

shiny and **bright**

Symmetric Patterns (SPs)

- Words that co-occur in SPs tend to be semantically **similar**
  - Widdows and Dorow, 2002
  - Davidov and Rappoport, 2006
  - Kozareva et al., 2008
  - Feng et al., 2013
  - **Schwartz** et al., 2014

Symmetric Patterns (SPs)

- Words that co-occur in SPs tend to be semantically **similar**
  - Widdows and Dorow, 2002
  - Davidov and Rappoport, 2006
  - Kozareva et al., 2008
  - Feng et al., 2013
  - **Schwartz** et al., 2014

John *and* Mike

neither here *nor* there

bold *and* beautiful

Paris *or* Rome

Symmetric Patterns (SPs)

- Words that co-occur in SPs tend to be semantically **similar**
  – Widdows and Dorow, 2002
  – Davidov and Rappoport, 2006
  – Kozareva et al., 2008
  – Feng et al., 2013
  – **Schwartz** et al., 2014

neither here *nor* there

#car *or* wheel

John *and* Mike

#*neither* cup *nor* coffee

bold *and* beautiful

Paris *or* Rome

#dog *and* leash

SP-based Word Embeddings

$$V^{sp}_{\mathbf{dog}} = \begin{pmatrix} \text{PPMI(dog,house)} \\ \text{PPMI(dog,mouse)} \\ \text{PPMI(dog,zebra)} \\ \text{PPMI(dog,wine)} \\ \text{PPMI(dog,cat)} \\ \text{PPMI(dog,dolphin)} \\ \text{PPMI(dog,bottle)} \\ \text{PPMI(dog,pen)} \\ . \\ . \\ . \end{pmatrix}$$

* Simple smoothing applied

SP-based Word Embeddings

$$V^{sp}_{\mathbf{dog}} = \begin{pmatrix} \text{PPMI(dog,house)} \\ \text{PPMI(dog,mouse)} \\ \text{PPMI(dog,zebra)} \\ \text{PPMI(dog,wine)} \\ \text{PPMI(dog,cat)} \\ \text{PPMI(dog,dolphin)} \\ \text{PPMI(dog,bottle)} \\ \text{PPMI(dog,pen)} \\ . \\ . \\ . \end{pmatrix}$$

✓ **similarity rather than relatedness**

* Simple smoothing applied

Antonyms

big / small

- Some SPs are indicative of antonymy (Lin et al., 2003)
  - “*either* X *or* Y” (*either* big *or* small)
  - “*from* X *to* Y” (*from* poverty *to* richness)

Antonyms

big / small

See Paper for Details!

- Some SPs are indicative of antonymy (Lin et al., 2003)
  - *"either X or Y"* (*either* big *or* small)
  - *"from X to Y"* (*from* poverty *to* richness)

Word Embeddings that Identify Antonyms

ACL 2015 Papers

- *Revisiting Word Embedding for Contrasting Meaning* (Chen et al.)
- *Learning Semantic Word Embeddings based on Ordinal Knowledge Constraints* (Liu et al.)
- *AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Lexemes* (Rothe and Schutze, **Best paper award**)

Word Embeddings that Identify Antonyms

ACL 2015 Papers

- *Revisiting Word Embedding for Contrasting Meaning* (Chen et al.)
- *Learning Semantic Word Embeddings based on Ordinal Knowledge Constraints* (Liu et al.)
- *AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Lexemes* (Rothe and Schutze, **Best paper award**)

First model to support for antonyms without using a dictionary or a thesaurus!

SP-based Word Embeddings

$$V^{sp}_{\mathbf{dog}} = \begin{pmatrix} \text{PPMI(dog,house)} \\ \text{PPMI(dog,mouse)} \\ \text{PPMI(dog,zebra)} \\ \text{PPMI(dog,wine)} \\ \text{PPMI(dog,cat)} \\ \text{PPMI(dog,dolphin)} \\ \text{PPMI(dog,bottle)} \\ \text{PPMI(dog,pen)} \\ . \\ . \\ . \end{pmatrix}$$

✓ **similarity rather than relatedness**

* Simple smoothing applied

SP-based Word Embeddings

$$V^{sp}_{dog} = \begin{pmatrix} PPMI(dog,house) \\ PPMI(dog,mouse) \\ PPMI(dog,zebra) \\ PPMI(dog,wine) \\ PPMI(dog,cat) \\ PPMI(dog,dolphin) \\ PPMI(dog,bottle) \\ PPMI(dog,pen) \\ . \\ . \\ . \end{pmatrix}$$

✓ **similarity rather than relatedness**

✓ **support for antonyms**

* Simple smoothing applied

Experiments

- Embeddings are generated using an 8G words corpus

- Baselines: six state-of-the-art models

- Word similarity task
  - SimLex999 dataset (Hill et al., 2014)

Results

Model	Spearman's ρ
Glove (Pennington et al., 2014)	0.35
PPMI-Bag-of-words	0.423
word2vec CBOW (Mikolov et al,. 2013)	0.43
Dep (Levy and Goldberg, 2014)	0.436
NNSE (Murphy et al., 2012)	0.455
word2vec skip-gram (Mikolov et al,. 2013)	0.462

Results

Model	Spearman's ρ
Glove (Pennington et al., 2014)	0.35
PPMI-Bag-of-words	0.423
word2vec CBOW (Mikolov et al,. 2013)	0.43
Dep (Levy and Goldberg, 2014)	0.436
NNSE (Murphy et al., 2012)	0.455
word2vec skip-gram (Mikolov et al,. 2013)	0.462
SP	**0.517**

Results

Model	Spearman's ρ
Glove (Pennington et al., 2014)	0.35
PPMI-Bag-of-words	0.423
word2vec CBOW (Mikolov et al,. 2013)	0.43
Dep (Levy and Goldberg, 2014)	0.436
NNSE (Murphy et al., 2012)	0.455
word2vec skip-gram (Mikolov et al,. 2013)	0.462
SP	**0.517**

$$f_{joint}(w_i, w_j) = \alpha \cdot f_{SP}(w_i, w_j) + (1-\alpha) \cdot f_{skip-gram}(w_i, w_j)$$

Results

Model	Spearman's ρ
Glove (Pennington et al., 2014)	0.35
PPMI-Bag-of-words	0.423
word2vec CBOW (Mikolov et al,. 2013)	0.43
Dep (Levy and Goldberg, 2014)	0.436
NNSE (Murphy et al., 2012)	0.455
word2vec skip-gram (Mikolov et al,. 2013)	0.462
SP	**0.517**
Joint	**0.563**

$$f_{joint}(w_i, w_j) = \alpha \cdot f_{SP}(w_i, w_j) + (1-\alpha) \cdot f_{skip-gram}(w_i, w_j)$$

Results

Model	Spearman's ρ
Glove (Pennington et al., 2014)	0.35
PPMI-Bag-of-words	0.423
word2vec CBOW (Mikolov et al,. 2013)	0.43
Dep (Levy and Goldberg, 2014)	0.436
NNSE (Murphy et al., 2012)	0.455
word2vec skip-gram (Mikolov et al,. 2013)	0.462
SP	**0.517**
Joint	**0.563**

$$f_{joint}(w_i, w_j) = \alpha \cdot f_{SP}(w_i, w_j) + (1-\alpha) \cdot f_{skip-gram}(w_i, w_j)$$

POS Analysis

Model	Adj.
Glove (Pennington et al., 2014)	0.571
PPMI-Bag-of-words	0.548
word2vec CBOW (Mikolov et al,. 2013)	0.579
Dep (Levy and Goldberg, 2014)	0.54
NNSE (Murphy et al., 2012)	0.594
word2vec skip-gram (Mikolov et al,. 2013)	0.604
SP	**0.663**

POS Analysis

Model	Adj.	Nouns
Glove (Pennington et al., 2014)	0.571	0.377
PPMI-Bag-of-words	0.548	0.451
word2vec CBOW (Mikolov et al,. 2013)	0.579	0.48
Dep (Levy and Goldberg, 2014)	0.54	0.449
NNSE (Murphy et al., 2012)	0.594	0.487
word2vec skip-gram (Mikolov et al,. 2013)	0.604	**0.501**
SP	**0.663**	0.497

POS Analysis

Model	Adj.	Nouns	Verbs
Glove (Pennington et al., 2014)	0.571	0.377	0.163
PPMI-Bag-of-words	0.548	0.451	0.276
word2vec CBOW (Mikolov et al,. 2013)	0.579	0.48	0.252
Dep (Levy and Goldberg, 2014)	0.54	0.449	0.376
NNSE (Murphy et al., 2012)	0.594	0.487	0.318
word2vec skip-gram (Mikolov et al,. 2013)	0.604	**0.501**	0.307
SP	**0.663**	0.497	**0.578**

POS Analysis

Model	Adj.	Nouns	Verbs
Glove (Pennington et al., 2014)	0.571	0.377	0.163
PPMI-Bag-of-words	0.548	0.451	0.276
word2vec CBOW (Mikolov et al,. 2013)	0.579	0.48	0.252
Dep (Levy and Goldberg, 2014)	0.54	0.449	0.376
NNSE (Murphy et al., 2012)	0.594	0.487	0.318
word2vec skip-gram (Mikolov et al,. 2013)	0.604	**0.501**	$\boxed{0.307}$
SP	**0.663**	0.497	$\boxed{\mathbf{0.578}}$

POS Analysis

Model	Adj.	Nouns	Verbs
Glove (Pennington et al., 2014)	0.571	0.377	$\boxed{0.163}$
PPMI-Bag-of-words	0.548	0.451	0.276
word2vec CBOW (Mikolov et al,. 2013)	0.579	0.48	0.252
Dep (Levy and Goldberg, 2014)	0.54	0.449	0.376
NNSE (Murphy et al., 2012)	0.594	0.487	0.318
word2vec skip-gram (Mikolov et al,. 2013)	0.604	**0.501**	0.307
SP	**0.663**	0.497	$\boxed{\mathbf{0.578}}$

More Results

- List of SPs is acquired automatically (**not manually defined**)
- Antonymy as Word Analogy
- Wordsim353 experiments
- And more…

More Results

- List of SPs is acquired automatically (**not manually defined**)
- Antonymy as Word Analogy
- Wordsim353 experiments
- And more…

See Paper!

Summary

- Word embeddings based on symmetric patterns
  - They capture **similarity** and not **relatedness**
  - The first word embeddings model to mark **antonym** pairs as dissimilar (w/o using a dictionary)

- Experiments on SimLex999
  - ***5.5%*** *improvement over six state-of-the-art models*
  - ***10%*** *improvement with a* ***joint*** *model*
  - ***20%*** *improvement on* ***verbs***

Future Work

- Enhancing bag-of-words models with SPs
- Does order count? **asymmetric** symmetric patterns

Roy Schwartz (*roys02@cs.huji.ac.il*)

ww.cs.huji.ac.il/~roys02/papers/sp_embeddings/sp_embeddings.html